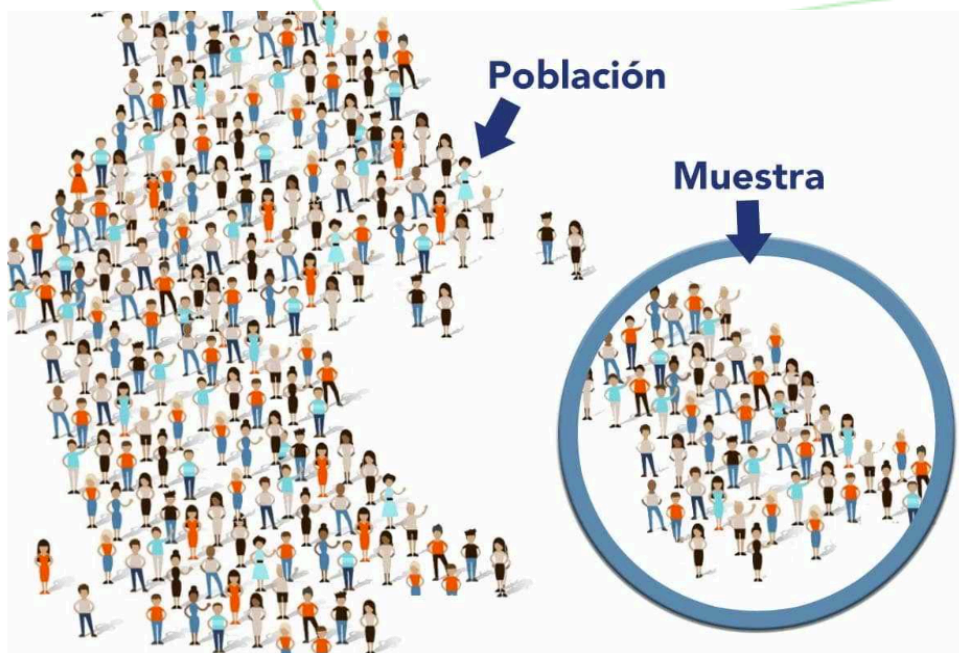


Estadística para las ciencias sociales

Semana 5

Poblaciones y muestras

En el campo de la estadística, especialmente cuando se aplica a las ciencias sociales, los conceptos población y muestra son fundamentales y forman la base para el diseño de estudios y la interpretación de datos. Su adecuada comprensión y aplicación es crucial pues determina la validez, fiabilidad y generalización de los resultados de cualquier estudio estadístico.



Nota: Imagen tomada de QuestionPro (s/f). **Tamaño de muestra.**

Población

La población constituye el objeto de la investigación, es su centro y de ella se extraerá la información requerida para su respectivo estudio. La población representa el conjunto de todas las unidades (personas o cosas) que concuerdan con una serie de especificaciones.

Existen dos consideraciones importantes que conviene tener presente con respecto al tamaño de la población, si es **población finita** o si es **población infinita**. En la población finita se conoce el número total de unidades que lo conforman, mientras que en la población infinita no es posible señalar con exactitud la cifra total de unidades.

Esta distinción es importante en la determinación del tamaño de la muestra, ya que para cada una de ellas existen fórmulas estadísticas que determinan el tamaño de la muestra, conservando un cierto nivel de confiabilidad. Cabe señalar que cuando el tamaño de la población es muy pequeño, se sugiere utilizar un diseño censal, es decir, trabajar con la población completa; en

este caso no se extrae una muestra, ya que se trabaja con la población total.

Muestra

Si la población hace referencias al conjunto completo de individuos, eventos o elementos que son el objeto de estudio, la muestra es un subconjunto de esta población, seleccionado para representarla en una investigación. Dicho subgrupo debe ser representativo de la población, que se extrae cuando no es posible medir cada una de las unidades de la población.

Las muestras se categorizan en dos grandes ramas: las **muestras no probabilísticas** y las **muestras probabilísticas**. En las muestras probabilísticas, todos los elementos de la población tienen la misma posibilidad de ser escogidos para la muestra, pues se obtienen definiendo las características de la población y el tamaño de la muestra, y por medio de una selección aleatoria o mecánica de las unidades de muestreo/análisis. En las muestras no probabilísticas, la elección de los elementos no depende de la probabilidad, sino de causas relacionadas con las características de la investigación o los propósitos del investigador.

Elegir entre una muestra probabilística o una no probabilística depende del planteamiento del estudio, del diseño de investigación y de la contribución que se piensa hacer con ella.

Tipos de muestreo

La elección de muestras implica determinar el método de muestreo y establecer el tamaño de la muestra, es decir, el número de elementos seleccionados. Revisemos el primero:

Método de muestreo	
Probabilístico	No probabilístico o intencional
Muestreo aleatorio simple	Muestreo por conveniencia
Muestreo aleatorio sistemático	Muestreo de juicio
Muestreo estratificado	Muestreo por cuotas
Muestreo por conglomerado	

Muestreo aleatorio simple

Se caracteriza porque cada elemento de la población tiene la misma probabilidad de ser elegido. Si se dispone de una lista de los elementos de la población, la selección de una muestra aleatoria simple es muy sencilla. Supongamos que numeramos los elementos de la población de 1 a N , una muestra aleatoria simple puede obtenerse colocando N bolas numeradas en una caja, mezclándolas y extrayendo n de ella.

Muestreo aleatorio sistemático

El muestreo sistemático es un método de selección de muestras empleado cuando se dispone de una lista ordenada de los elementos de la población, este orden no introduce un sesgo en el estudio. El proceso comienza seleccionando un primer elemento de manera aleatoria entre 1 y p , donde p es el cociente de dividir el tamaño total de la población (N) por el tamaño de la muestra deseada (n). A partir de este elemento inicial (k -ésimo), los siguientes se eligen sumando p al anterior ($k, k+p, k+2p, \dots$).

Este método tiende a generar muestras más representativas que el muestreo aleatorio simple, ya que asegura la inclusión de elementos de todas las clases de la población. Sin embargo, presenta el riesgo de introducir homogeneidad no representativa si existe periodicidad en la población, como en el caso de listas donde hay una distribución regular de ciertas características (por ejemplo, género o estacionalidad). En tales situaciones, el muestreo sistemático podría resultar en una muestra sesgada, al seleccionar elementos con características similares de manera constante.

Muestreo estratificado

Hace referencia a poblaciones con características diferentes, en las que existen estratos (grupos) homogéneos con respecto al carácter que se investiga y que, por tanto, deben ser analizados de forma diferente, ya que pueden obtenerse estimaciones más precisas (menos errores) tomando una muestra aleatoria en cada estrato. Además, se garantiza que todos los estratos (subpoblaciones, en definitiva) estén representados. Cada estrato tiene elementos que son homogéneos entre sí y heterogéneos con respecto a los demás estratos. Por ejemplo, si se quieren estudiar características de los hoteles de una determinada provincia, se tienen como estratos los hoteles de

5 estrellas, de 4 estrellas, de 3 estrellas, etc., y se utilizará un muestreo estratificado que nos asegure una adecuada representación de todos los estratos.

Muestreo por conglomerado

En el muestreo por conglomerados la unidad muestral es un grupo de elementos de la población que forman una unidad, a la que llamamos conglomerado (familias, empresas, municipios, etc.). El muestreo por conglomerados consiste en seleccionar aleatoriamente un cierto número de conglomerados (el necesario para alcanzar el tamaño muestral establecido) e investigar después todos los elementos pertenecientes a los conglomerados elegidos. En el caso particular de que los conglomerados sean áreas geográficas, bastante frecuente, hablaremos de muestreo por áreas.

Tabla 1: Ventajas e inconvenientes de los diferentes tipos de muestreo aleatorio

	CARACTERÍSTICAS	VENTAJAS	INCONVENIENTES
Aleatorio simple	Se selecciona una muestra de tamaño n de una población de N unidades, cada elemento tiene una probabilidad de inclusión igual y conocida de n/N .	Sencillo y de fácil comprensión.	Requiere que se posea de antemano un listado completo de toda la población. Caro.
Sistemático	Se necesita un listado de los N elementos de la población. Tras determinar el tamaño muestral n , se define un valor $p = N/n$. Se elige un número aleatorio, k , entre 1 y p ($k =$ arranque aleatorio) y se seleccionan los elementos de la lista que ocupan los lugares $k, k+p, k+2p, \dots$	Fácil de aplicar. Cuando la población está ordenada siguiendo una tendencia conocida, asegura una cobertura de unidades de todos los tipos.	Si la constante de muestreo está asociada con el fenómeno de interés, las estimaciones obtenidas a partir de la muestra pueden contener sesgo de selección
Estratificado	En ciertas ocasiones resultará conveniente estratificar la muestra según ciertas variables de interés. Para ello debemos conocer la composición estratificada de la población.	Tiende a asegurar que la muestra represente adecuadamente a la población en función de las variables seleccionadas. Estimaciones más precisas. Su objetivo es conseguir una muestra lo más semejante posible a la población en lo referente a las variables estratificadoras.	Se ha de conocer la distribución en la población de las variables utilizadas para la estratificación.
Conglomerados-Etapas	Se seleccionan aleatoriamente cierto número de conglomerados y se investigan todos los elementos de cada uno de ellos. El muestreo por etapas es una generalización, que va de más grandes a más pequeños. Se realizan varias fases de muestreo sucesivas.	Es muy eficiente cuando la población es muy grande y dispersa. No es preciso tener un listado de toda la población, sólo de las unidades últimas de muestreo. Menor coste.	El error estándar es mayor que en el muestreo aleatorio simple o estratificado. El cálculo del error estándar es complejo.

Nota: Tabla tomada de Morillas, A. (2010). *Muestreo en poblaciones finitas*.

Muestreo por conveniencia

El muestreo por conveniencia es un método de muestreo no probabilístico o intencional en el cual la selección de los elementos de la muestra se basa en la facilidad de acceso o disponibilidad, en lugar de en un proceso aleatorio. En este enfoque, el investigador elige los sujetos o elementos que son más accesibles o convenientes para formar parte del estudio. Dado que la selección no se realiza al azar, este método no garantiza que la muestra sea representativa de la población en su conjunto, lo que puede llevar a sesgos en los resultados.

El muestreo por conveniencia es comúnmente utilizado en investigaciones exploratorias o preliminares, estudios piloto, o cuando la representatividad no es una prioridad o los recursos son limitados. Sin embargo, debido a su naturaleza no aleatoria, los hallazgos obtenidos a través de este tipo de muestreo deben interpretarse con cautela y no se consideran generalizables a toda la población.

Muestreo de juicio

Este tipo de muestreo se caracteriza por un esfuerzo deliberado de obtener muestras "representativas" mediante la inclusión en la muestra de grupos supuestamente típicos. Para la obtención de una muestra mediante este método se puede acudir a criterios específicos de los investigadores o bien requerir las opiniones de expertos en la materia, por lo que también es denominado como muestreo de opinión. Por ejemplo, cuando el interés del estudio se centra en comparar las características diferenciadoras de los clientes que han presentado cierta insatisfacción en el servicio con respecto a aquellos otros que no, se determina a juicio del investigador un número predeterminado de encuestas a los clientes satisfechos y otro número de encuestas a clientes insatisfechos.

Muestreo por cuotas

También denominado en ocasiones "accidental". Se asienta generalmente sobre la base de un buen conocimiento de los estratos de la población y/o de los individuos más "representativos" o "adecuados" para los fines de la investigación. Mantiene semejanzas con el muestreo aleatorio estratificado, pero no tiene el carácter de aleatoriedad de aquél. En este tipo de muestreo

se fijan unas "cuotas" que consisten en un número de individuos que reúnen determinadas condiciones (variables de control), por ejemplo: 20 individuos de 25 a 40 años, de sexo femenino y residentes en Málaga. Siempre que se ajuste a las cuotas fijadas, el entrevistador tiene libertad para elegir a los entrevistados. Este método se utiliza mucho en las encuestas de opinión.

Error de muestreo

En estadística, error muestral o error de estimación es el error a causa de observar una muestra en lugar de la población completa.

La estimación de un valor de interés, como la media o el porcentaje, estará generalmente sujeta a una variación entre una muestra y otra. Estas variaciones en las posibles muestras de una estadística pueden, teóricamente, ser expresadas como errores muestrales; sin embargo, por lo regular, en la práctica el error exacto es desconocido. El error muestral se refiere en términos más generales al fenómeno de la variación entre muestras. Cuando este no es mencionado se considera que el margen de error base es el 0.02% (0.2 para muestreo paralelo y 2 para muestreo directo).

El error muestral deseado, en general puede ser controlado tomando una muestra aleatoria suficientemente grande de la población; sin embargo, el costo de esto puede ser limitante. Si las observaciones son tomadas de una muestra aleatoria, la teoría estadística brinda cálculos probabilísticos del tamaño deseado del error muestral para una estadística en particular o estimación. Estos usualmente son expresados en términos del error estándar.

El error muestral puede ser contrastado con el error no muestral, el cual se refiere al conjunto de las desviaciones del valor real que no van en función de la muestra escogida, entre los cuales se encuentran varios errores sistemáticos y algunos errores aleatorios. Resultan mucho más difíciles de cuantificar que el error muestral.

Tamaño de muestra

El tamaño de muestra hace referencia al número de individuos, casos u observaciones que se incluyen en una muestra para un estudio. Es una consideración crucial en la planificación de investigaciones, ya que influye en la capacidad de un estudio para detectar efectos o tendencias significativas, así como en la precisión y fiabilidad de los resultados estadísticos.

El cálculo del tamaño de la muestra se realiza mediante dos fórmulas, según sea la **población: finita o infinita**, para obtener las fórmulas que permitirán determinar el tamaño de la muestra se sugiere seguir los siguientes pasos:

1. Determinar el grado de confianza y el valor de Z para el grado de confiabilidad ya seleccionado. Para establecer el valor de Z, según el margen de confianza, Navarro (2009, en Población y muestra, s/f) elaboró una tabla que podría facilitar el entendimiento y cálculo de la fórmula para la determinación de las muestras finitas o infinitas:

PASO 1	
GRADO DE CONFIANZA	VALOR DE Z PARA DIFERENTES GRADOS DE CONFIABILIDAD
80%	1,29
85%	1,44
90%	1,64
95%	1,96
99%	2,57

Nota: Tabla tomada de **Población y muestra**.

2. Se procede a determinar la probabilidad de ocurrencia de evento favorable y probabilidad de ocurrencia de evento desfavorable, que por lo general y a juicio de Navarro (2009, en Población y muestra, s/f) “cuando no se tiene una idea clara de esta situación, es necesario dar sus máximos valores, tanto a la probabilidad de que se realice el evento favorable, como a la que no se realice” (p. 2), es decir, 50% de probabilidad de ocurrencia favorable del evento (p) y 50% probabilidad de ocurrencia de evento desfavorable (q).

PROBABILIDAD DE OCURRENCIA FAVORABLE O DESFAVORABLE DEL EVENTO A ESTUDIAR			
REALIZACIÓN DEL EVENTO		NO REALIZACIÓN DEL EVENTO	
%	LITERAL EMPLEADO	%	LITERAL EMPLEADO
50	p	50	q

Nota: Tabla tomada de **Población y muestra**.

3. Se establece el error máximo que puede ser aceptado en los resultados; por lo regular, manifiesta Navarro (2009) “se trabaja hasta con un 6% de error, ya que las variaciones superiores al 10% reducirán demasiado la validez de la información” (p. 57), y en opinión de la autora de la presente guía, lo recomendable sería trabajar con un 95% de confianza y un margen de error de 5%.

Cálculo del tamaño muestral en poblaciones infinitas

En situaciones donde la población es muy grande se considera infinita para fines prácticos, el tamaño de la muestra se calcula sin necesidad de ajustar por el tamaño de la población. En estos casos, el cálculo se enfoca principalmente en el nivel de confianza y el margen de error, así como en la variabilidad esperada de los datos.

$$n = \frac{Z^2 \cdot p \cdot q}{e^2}$$

Donde:

n= Tamaño de la muestra

Z= Coeficiente de confianza

p= Probabilidad a favor

q= Probabilidad en contra

e= Error de estimación

Cálculo del tamaño muestral en poblaciones finitas

La población finita hace referencia a poblaciones donde el número total de sujetos es conocido y limitado. En el caso de poblaciones finitas, el tamaño de la muestra se calcula teniendo en cuenta el tamaño total de la población.

$$n = \frac{Z^2 \cdot N \cdot p \cdot q}{e^2(N - 1) + Z^2 \cdot p \cdot q}$$

Donde:

Z= Coeficiente de confianza

N= Población

p= Probabilidad a favor

q= Probabilidad en contra

e= Error de estimación

n= Tamaño de la muestra

Cualquiera de estas dos fórmulas puede emplearse, sólo que hay que tener presente que la segunda fórmula dará la mayor parte de las veces un tamaño de muestra más grande.

Un muestreo bien planificado y ejecutado puede proporcionar *insights* significativos sobre una población más amplia, permitiendo a los investigadores hacer inferencias y conclusiones con un nivel de confianza

razonable. Sin embargo, desafíos como el sesgo de selección y la representatividad de la muestra pueden tener un impacto profundo en la interpretación de los datos y, por ende, en las conclusiones extraídas. Por lo tanto, entender la relación entre población y muestra, y cómo manejar adecuadamente estos conceptos, es esencial para cualquier investigación en el ámbito de las ciencias sociales.

Referencias

Población y muestra (s/f).

https://apps.utel.edu.mx/recursos/files/r161r/w26286w/GUIA_POBLACION_Y_MUESTRA.pdf

QuestionPro (s/f). **Tamaño de muestra.**

<https://www.questionpro.com/es/tama%C3%B1o-de-la-muestra.html>

Morillas, A. (2010). **Muestreo en poblaciones finitas.**

https://www.u-cursos.cl/ingenieria/2010/1/IN3401/1/material_docente/bajar?id_material=280296

Morillas, A. (2010). **Muestreo en poblaciones finitas.**

https://www.u-cursos.cl/ingenieria/2010/1/IN3401/1/material_docente/bajar?id_material=280296

Métodos cuantitativos de investigación (s/f). **Muestra estadística, Muestreo en poblaciones finitas, Error muestral, Estrategias de muestreo y Muestreo estratificado.**

<https://cursos.aiu.edu/METODOS%20CUANTITATIVOS%20DE%20INVESTIGACION/8/Sesi%C3%B3n%208.pdf>